Research on Frequency Estimation Based on LS-SVC in Unknown Noise

Xueqian Liu
Zhengzhou Information Science and Technology Institute, Zhengzhou 450002, China
Email: jxdlxq@126.com

Abstract—When noise model is already known, maximum likelihood estimator (MLE) is asymptotically the most optimum one. However, the truth is just the opposite that, noise is unknown in burst communication systems. Aiming that, this paper utilizes the transmitted symbols of burst communications which are eliminated firstly in common methods and proposes a data-aided (DA) frequency estimation algorithm based on least squares support vector classification (LS-SVC). By researching on statistical learning theory (SLT), we construct a structural risk minimization (SRM) function with respect to frequency, and convert the estimation problem into deriving the extremum value of a classification function. Consequently, support vector classification (SVC)'s good learning and generalization capabilities are completely explored and employed. Experimental results show that the proposed algorithm is close to MLE in the case of Gaussian noise, and also exhits good performance in non-Gaussian condition.

Index Terms—frequency estimation, least squares support vector classification (LS-SVC), unknown noise, gaussian distribution, α -stable distribution, support vector classification (SVC), structural risk minimization (SRM)

I. INTRODUCTION

Frequency estimation of sinusoidal signal with noise is a classical problem in signal processing, which has started from 1970s [1]-[8]. And we usually suppose that the noise models are Gaussian distribution ones. However, as a matter of fact, the assumption of that is impractical in many scenes. Resultly, frequency estimation algorithms in the condition of non-Gaussian noise are addressed [9]-[13]. Even though with that being the case, their distributions are always known, such as α -stable and pulse ones. Differently, frequency estimation under the assumption of unknown noise distribution is discussed in this paper. At this moment, the traditional methods based on a priori knowledge are invalid. And we utilize the machine learning method to obtain model information as much as possible.

The relationships between inputs and outputs are assumed already known in artificial neural network (ANN), only with unknown parameters. Hence, over-matching and local minimum problems always exist. Distinctly, statistical learning theory (SLT) which is specialized in the research of small sample condition, tries to get inner connections only by independent and identically distributed (i.i.d.) data samples [14]. As its concrete implement, support vector classification (SVC) exhibits good performances in generalizing, high-dimensional processing and nonlinear processing. Where least squares support vector classification (LS-SVC) has the improvements: inequality constraints are substituted by equality ones; a squared loss function is taken for the error variable.

The rest of this paper is organized as follows. Section II briefly introduces the basic theory of SLT and LS-SVC. Frequency estimation algorithm based on LS-SVC is proposed in Section III, and the proper choices of LS-SVC's parameters are also discussed, taking binary phase shift keying (BPSK) signal for example. Then in Section IV, simulations and experiments verifies the feasibility and validity of the proposed algorithm. At last, conclusions of the paper are given in Section V.

II. SLT AND LS-SVC

A. SLT

Aiming at binary classification, the joint probability density function (PDF) of set $S=\{(\mathbf{x}_i, y_i) | i=1,\cdots,N\}$ is unknown, which is denoted as $P(\mathbf{x}, y)$, where $\mathbf{x}_i \in \mathbb{R}^n$ is n dimension characteristic vector, $y_i \in \{-1,1\}$ is category label. Now, we try to select a optimizing one from the response set of learning machine $f(\mathbf{x},\mathbf{w})$ to minimize the expected risk, i.e.:

$$R(\mathbf{w}) = \int L(y, f(\mathbf{x},\mathbf{w}))\, dP(\mathbf{x}, y) \quad (1)$$

where $L(y, f(\mathbf{x},\mathbf{w}))$ is the loss function used to show the difference between response of training machine y and learning machine $f(\mathbf{x},\mathbf{w})$. According to binary classification, it is denoted as

$$L(y, f(\mathbf{x},\mathbf{w})) = \begin{cases} 0 & \text{if } y = f(\mathbf{x},\mathbf{w}) \\ 1 & \text{if } y \neq f(\mathbf{x},\mathbf{w}) \end{cases}$$

Manuscript received July 17, 2013, revised October 28, 2013.
This work was supported by the National Nature Science Foundation of China under Grant No. 61201380.
Corresponding author email: jxdlxq@126.com
doi: 10.12720/jcm.8.10.643-650

where $\mathbf{w}$ is the argument of $f(\mathbf{x},\mathbf{w})$. And, the different type of learning problem has the different $L(y, f(\mathbf{x},\mathbf{w}))$ which can undoubtly impact the its learning performance.

It is proved that if η is a random number during the range $[0,1]$, the inequation is satisfied as follows, with probability $1-\eta$ [14]:

$$R(\mathbf{w}) \le R_{\text{emp}}(\mathbf{w}) + \Phi(h/N) = R_{\text{emp}}(\mathbf{w}) + \sqrt{\frac{h(\ln(2N/h)+1) - \ln(\eta/4)}{N}} \quad (2)$$

where $R_{\text{emp}}(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i,\mathbf{w}))$ is empirical risk function, $\Phi(h/N)$ is confidence limit and determines its generalization capability, h is Vapnik-Chernonekis (VC) dimension and denotes its complexity.

As $R_{\text{emp}}(\mathbf{w})$ is directly proportional to h, $\Phi(h/N)$ is inversely proportional to h, structural risk minimiz-etion (SRM) rule is presented to minimizing $R(\mathbf{w})$. It has divided $S = \{f(x,w)\}$ into a series of subset:

$$S_1 \subset S_2 \subset \ldots \subset S_k \subset \ldots \subset S; S = \cup S_i \quad (3)$$

where their VC dimension is arranged by:

$$h_1 \le h_2 \le \ldots \le h_k \le \ldots \quad (4)$$

In each subset, the rule tries to find the function having the smallest empirical risk. At the same time, it considers the compromise between empirical risk and confidence limit, and ultimately derive the smallest expected risk, which is depicted in Fig. 1.

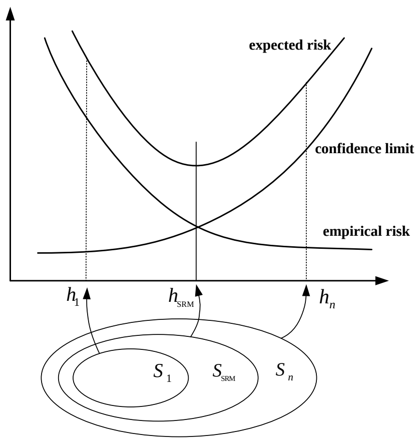


Figure 1. Structure risk minimization

B. LS-SVC

Supposed that, S can be divided by a hyperplane $(\mathbf{w}\cdot\phi(\mathbf{x}))+b=0$ without errors, where $(\cdot)$ is an inner product operation and $\phi(\cdot)$ is a nonlinear mapping from low to high dimension characteristic space.

The hyperplane can be normalized, and the boundary samples are satisfied by:

$$\begin{aligned} (\mathbf{w}\cdot\phi(\mathbf{x}_i))+b=1 \quad &\text{if } y_i=1 \\ (\mathbf{w}\cdot\phi(\mathbf{x}_i))+b=-1 \quad &\text{if } y_i=-1 \end{aligned} \quad (5)$$

Maximizing the margin between two sorts which equals $2/\|\mathbf{w}\|$ and getting the optimal hyperplane:

$$\begin{aligned} &\min \ J(\mathbf{w},b) = \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{s.t.} \ y_i\left[(\mathbf{w}\cdot\phi(\mathbf{x}_i))+b\right]-1 \ge 0, \ i=1,\cdots,N \end{aligned} \quad (6)$$

By introuducing error variables e_i and LS method, (6) is converted into:

$$\begin{aligned} &\min \ J(\mathbf{w},b) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{2}\left(\sum_{i=1}^{N} e_i^2\right) \\ &\text{s.t.} \ y_i\left[(\mathbf{w}\cdot\phi(\mathbf{x}_i))+b\right] = 1-e_i, \ i=1,\cdots,N \end{aligned} \quad (7)$$

Minimizing the first item of $J(\mathbf{w},b)$ means that the margin of (5) now is the largest, and minimizing the second one means that both of the boundaries now are in the center of their sort as much as possible. Penalty factor C is a positive constant and can control the penalty degree of fitting errors. It means that, SRM is introduced into LS-SVC to make compromise in confidence limit and empirical risk.

Equation (7) is a strict convex quadratic programming (QP) problem in optimization theory. Using Lagrange multiplier method, where α_i is Lagrange multiplier:

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \phi(\mathbf{x}_i) \quad (8)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \quad (9)$$

$$\alpha_i = Ce_i \quad (10)$$

Combining (7), (8), (9) and (10), and replacing $(\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j))$ with kernel function $K(\mathbf{x}_i,\mathbf{x}_j)$:

$$\begin{bmatrix} 0 & \mathbf{Y}^{\mathrm{T}} \\ \mathbf{Y} & \mathbf{PQP}^{\mathrm{T}} + \frac{1}{C}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{E} \end{bmatrix} \quad (11)$$

where $\boldsymbol{\alpha} = (\alpha_1,\cdots,\alpha_N)^{\mathrm{T}}$, $\mathbf{E} = (1,\cdots,1)^{\mathrm{T}}$, $\mathbf{Y} = (y_1,\cdots,y_N)^{\mathrm{T}}$, $\mathbf{P}$ is a diagonal matrix whose main diagonal elements are $y_1,\cdots,y_N$, $\mathbf{Q}$ is named as kernel function matrix. Beacause we select radius basis function (RBF) in this study, so the (i,j)th element of $\mathbf{Q}$ is:

$$\mathbf{Q}_{ij} = K(\mathbf{x}_i,\mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{h^2}\right) \quad (12)$$

Ultimately, the discriminant function is obtained as

$$\begin{aligned} f(\mathbf{x}) &= \operatorname{sgn}\left[(\mathbf{w}\cdot\phi(\mathbf{x}))+b\right] \\ &= \operatorname{sgn}\left[\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i,\mathbf{x})+b\right] \end{aligned} \quad (13)$$

III. Frequency Estimation Algorithm Based on LS-SVC

A. Signal Model

The sinusoidal signal polluted by noise is modeled as

$$r_n = a_n A e^{j(2\pi f_0 n T_s + \theta_0)} + w_n, \quad n = 0, \cdots, N-1 \qquad (14)$$

where a_n is an independent symbol, M is the modulation order; $A > 0,\ f_0 \in [-0.5, 0.5), \theta_0 \in [-\pi, \pi)$ are amplitude, deterministic but unknown frequency and initial phase, respectively; T_S is the sample period, N is the sample size; w_n is an independent complex noise with zero-mean and unknown PDF. For the sake of simplicity, we firstly set $T_s = 1$ and take binary phase shift keying (BPSK) signal for example.

As a common non-Gaussian distribution, α -stable one only has unified characteristic function [15]:

$$\varphi(t) = \exp\left(j\mu t - \gamma |t|^{\alpha}\left[1 + j\beta \operatorname{sgn}(t) w(t, \alpha)\right]\right) \qquad (15)$$

where

$$w(t,\alpha) = \begin{cases} \tan(\alpha\pi/2) & \text{if } \alpha \neq 1 \\ (2/\pi)\ln|t| & \text{if } \alpha = 1 \end{cases},$$

where $\alpha \in [0, 2]$ is characteristic exponent and describes the thickness of tails, when $\alpha = 2$, it is Gaussian distribution; when $\alpha = 1, \beta = 0$, it is Cauchy distribution. $\beta \in [-1, 1]$ is skewness parameter, when $\beta = 0$, it is symmetrical about μ and called $S\alpha S$ for short. $\gamma \in (0, +\infty)$ is scale parameter and similar with the variance of Gaussian distribution. $\mu \in (-\infty, +\infty)$ are shift parameter, when $1 \le \alpha \le 2$, μ is the mean value; when $0 < \alpha < 1$, μ is the intermediate value. We set $\alpha = 1.5, \beta = 0, \gamma = 1, \mu = 0$ in this study, and plot the distribution of 1000 real samples in Fig. 2.

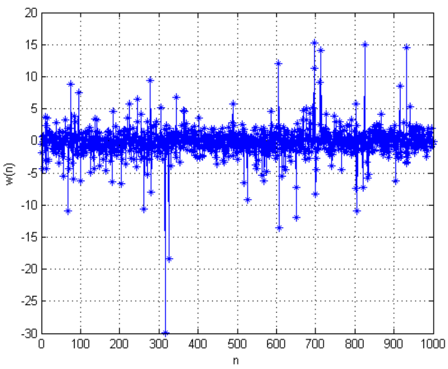


Figure 2. Real α -stable distribution with $\alpha = 1.5, \beta = \mu = 0, \gamma = 1$

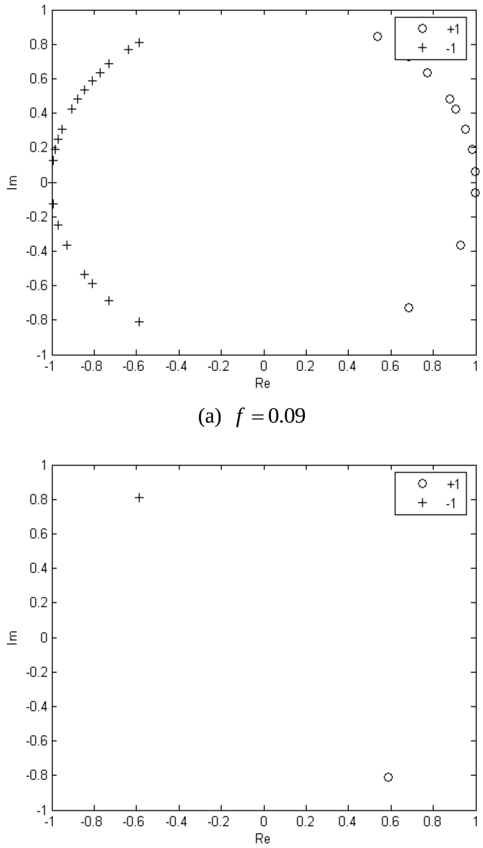
Contemporarily, no limited second-order moment is existing in fractional low-order α -stable distribution. Accordingly, its variance and power are both meaningless, and the gain of signal-to-noise ratio (GSNR) is defined as

$$10\lg\left(\frac{1}{2C_g}\left(\frac{A}{S_0}\right)^2\right)$$

where

$$S_0 = \exp\left(\frac{1}{N}\sum_{i=0}^{N-1}\ln|w_i|\right),$$

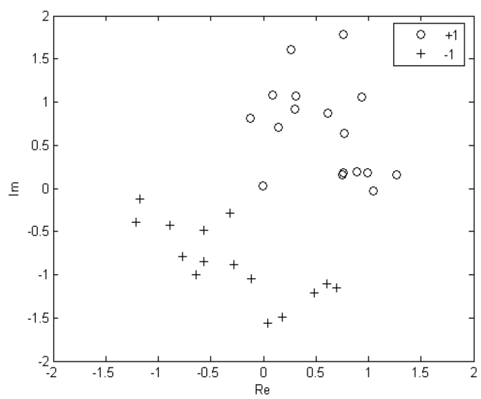
and $C_g = 1.78$ is a positive constant.



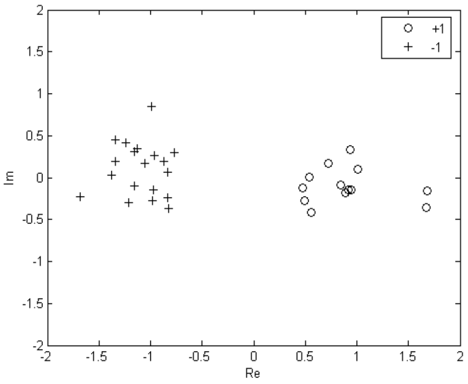
(a) $f = 0.09$

(b) $f = 0.1$

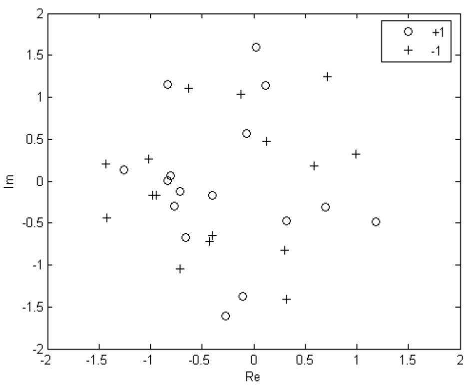
Figure 3. Constellations of $r_n^{'}$ without noise



(a) $f = 0.09$

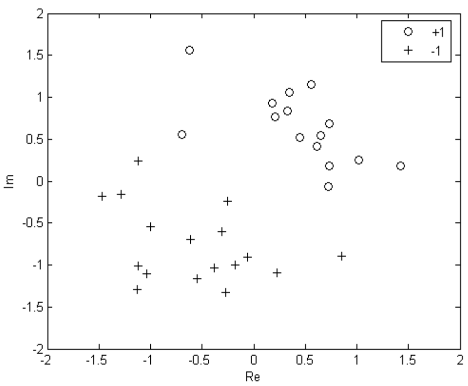


(b) $f = 0.1$

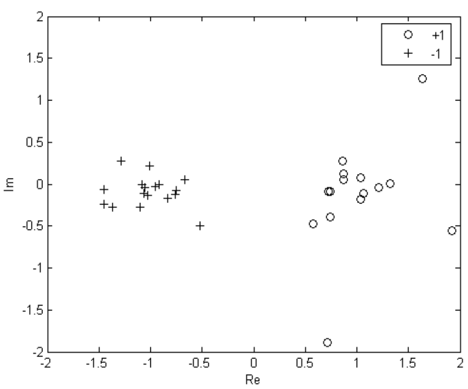


(c) $f = 0.13$

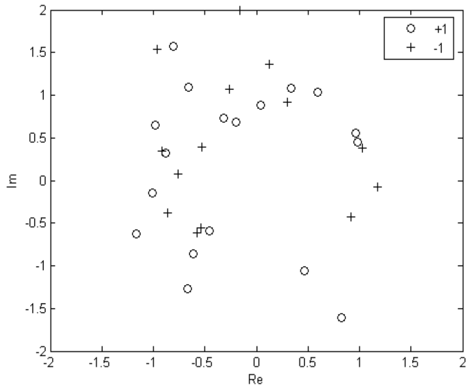
Figure 4. Constellations of $r_n^{'}$ when noise is Gaussian distribution (SNR is 6dB)



(a) $f = 0.09$



(b) $f = 0.1$



(c) $f = 0.13$

Figure 5. Constellations of $r_n^{'}$ when noise is α -stable distribution (GSNR is 6dB)

B. Frequency Estimation Algorithm Based on LS-SVC

We construct $b_n = e^{-j(2\pi fn+\theta)}$ after setting a frequency $f \in [-0.5, 0.5)$ and phase $\theta \in [-\pi, \pi)$. Letting

$$\begin{aligned} r_n^{'} &= r_n b_n = r_n e^{-j(2\pi fn+\theta)} \\ &= A a_n e^{j[2\pi(f_0 - f)n + (\theta_0 - \theta)]} + w_n e^{-j(2\pi fn+\theta)} \end{aligned} \tag{16}$$

(16) means, the influence of deterministic frequency f and initial phase θ are removed from received signal. We divide $r_n^{'}$ into two sorts by $a_n = +1$ or $a_n = -1$, i.e., we construct the training set $S = \{(\mathbf{x}_n(f), y_n) | n = 1, \cdots, N\}$, $\mathbf{x}_n(f) = [\mathrm{Nor}\{\mathrm{Re}[r_{n-1}^{'}(f)]\}, \mathrm{Nor}\{\mathrm{Im}[r_{n-1}^{'}(f)]\}]$, $y_n = a_{n-1}$, where $\mathrm{Nor}\{\cdot\}, \mathrm{Re}[\cdot], \mathrm{Im}[\cdot]$ are normalized, taking real and image part operators, respectively. We make use of LS-SVC and derive the SRM function with respect to f:

$$T(f) = \min_{\mathbf{w},b}[J(f, \mathbf{w}, b)] \tag{17}$$

Integrating (7) and (8), (17) is converted into:

$$\begin{aligned} T(f) = &\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}(f)\mathbf{PQ}(f)\mathbf{P}^{\mathrm{T}}\boldsymbol{\alpha}(f) \\ &+ \frac{1}{2C}\boldsymbol{\alpha}^{\mathrm{T}}(f)\boldsymbol{\alpha}(f) \end{aligned} \tag{18}$$

where $\boldsymbol{\alpha}(f)$ is the solution of (11).

Searching the hole interval of f, and getting the estimation value of frequency:

$$\hat{f} = \underset{f \in [-0.5, 0.5)}{\arg\min}\{T(f)\} \tag{19}$$

At first, we consider the unnoisy condition, now $r_n^{'} = A a_n e^{j[2\pi(f_0 - f)n + (\theta_0 - \theta)]}$. Setting $f_0 = 0.1, \theta_0 = 0, A = 1, N = 32, \theta = 0.5\pi$, the constellations of $r_n^{'}$ with different f are illustrated in Fig. 3. We can see that only if

$f = f_0$, $r_n^{'}$ are two discrete points; otherelse, $r_n^{'}$ is different with n. As a result, even if both two sorts are completely separable and the second item of (18) is equal to 0 with the different f, $T(0.1)$ is still the minimized value because of the first one. At the same time, $e^{-j\theta}$ means the clockwise rotation of a fixed angle θ in constellation. Therefore, the value of θ will not have an impact on the proposed algorithm. The experimental simulations in Part IV are consistent with this conclusion, so $\theta = 0$ is set in this study.

Then, we take the noisy condition into account. Setting $\theta = 0$, SNR or $\mathrm{GSNR} = 6\mathrm{dB}$, others are as in Fig. 3, Fig. 4 and Fig. 5 illustrate the constellations of $r_n^{'}$ with different f while noises are independent complex Gaussian distribution and α-stable distribution, respectively, where Gaussian distribution has zero-mean and variance σ^2, the parameter setting of α-stable distribution is the same as Fig. 2. From them, we know that whether the noise distribution is, $T(0.1)$ keeps the minimized one.

C. Parameter Settings of LS-SVC

From (11):

$$b = \frac{\mathbf{Y}^{\mathrm{T}}\left(\mathbf{PQP}^{\mathrm{T}} + \frac{1}{C}\mathbf{I}\right)^{-1}\mathbf{E}}{\mathbf{Y}^{\mathrm{T}}\left(\mathbf{PQP}^{\mathrm{T}} + \frac{1}{C}\mathbf{I}\right)^{-1}\mathbf{Y}} \tag{20}$$

$$\boldsymbol{\alpha} = \left(\mathbf{PQP}^{\mathrm{T}} + \frac{1}{C}\mathbf{I}\right)^{-1}\left(\mathbf{I} - \frac{\mathbf{YY}^{\mathrm{T}}\left(\mathbf{PQP}^{\mathrm{T}} + \frac{1}{C}\mathbf{I}\right)^{-1}}{\mathbf{Y}^{\mathrm{T}}\left(\mathbf{PQP}^{\mathrm{T}} + \frac{1}{C}\mathbf{I}\right)^{-1}\mathbf{Y}}\right)\mathbf{E} \tag{21}$$

When C decreases rapidly, $\mathbf{PQP}^{\mathrm{T}} + \frac{1}{C}\mathbf{I} \approx \frac{1}{C}\mathbf{I}$, then (21) and (19) are simplified as

$$\boldsymbol{\alpha}(f) \approx C\left(\mathbf{I} - \frac{\mathbf{YY}^{\mathrm{T}}}{\mathbf{Y}^{\mathrm{T}}\mathbf{Y}}\right)\mathbf{E} = C\left(\mathbf{I} - \frac{\mathbf{YY}^{\mathrm{T}}}{N}\right)\mathbf{E} \tag{22}$$

$$\hat{f} = \underset{f \in [-0.5, 0.5)}{\arg\min}\left\{\frac{1}{2C}\boldsymbol{\alpha}^{\mathrm{T}}(f)\boldsymbol{\alpha}(f)\right\} \tag{23}$$

From (22) and (23), we can know that $\hat{f}$ with different $f_0 - f$ are same and very small, for the reason of the equal $\boldsymbol{\alpha}(f)$ with different $f_0 - f$. As a result, the proposed algorithm fails. Summarizingly, we must select C as large as possible. However, too large C will lead that $\mathbf{PQP}^{\mathrm{T}} + \frac{1}{C}\mathbf{I} \approx \mathbf{PQP}^{\mathrm{T}}$, then (21) and (19) are simplified as

$$\boldsymbol{\alpha}(f) = \left(\mathbf{PQ}(f)\mathbf{P}^{\mathrm{T}}\right)^{-1}\left(\mathbf{I} - \frac{\mathbf{YY}^{\mathrm{T}}\left(\mathbf{PQ}(f)\mathbf{P}^{\mathrm{T}}\right)^{-1}}{\mathbf{Y}^{\mathrm{T}}\left(\mathbf{PQ}(f)\mathbf{P}^{\mathrm{T}}\right)^{-1}\mathbf{Y}}\right)\mathbf{E} \tag{24}$$

$$\hat{f} = \underset{[-0.5, 0.5]}{\arg\min}\left\{\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}(f)\mathbf{PQ}(f)\mathbf{P}^{\mathrm{T}}\boldsymbol{\alpha}(f)\right\} \tag{25}$$

As (24) and (25) shown, now C barely influences $\boldsymbol{\alpha}(f)$ and $\hat{f}$ which is only determined by $\mathbf{Q}(f)$. Accordingly, the estimation accuracy of the proposed algoriththm decreases. Everything is as in Fig. 5 other than setting the width of RBF $h = 1$, and the number of Monte Carlo experiments is 1000, Fig. 6 illustrates the mean square error (MSE) curves of the proposed algorithm with that C are 0.1, 1, 1000 and 10000, resectively. Fig. 6 is consistent with all above analyses, and we select $C = 1000$ in this study.

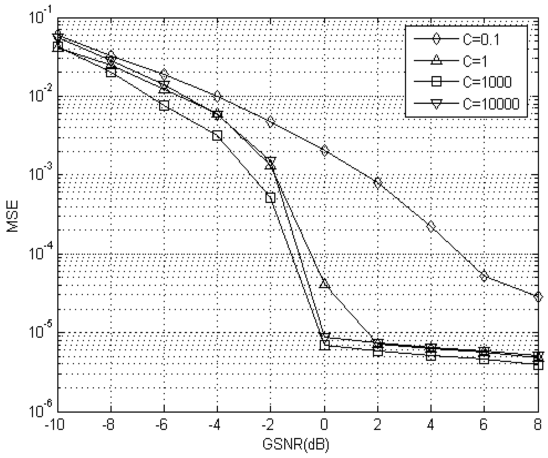


Figure 6. Impact of C on MSE

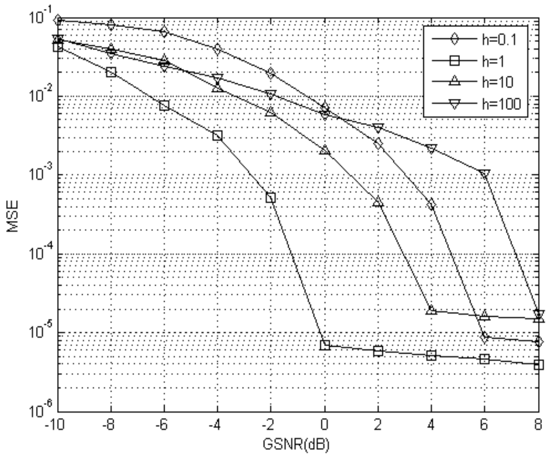


Figure 7. Impact of h on MSE

According to (12), the main diagonal elements of $\mathbf{Q}$ keep 1 all the time. When h decreases, other elements gradually approach to 0, thus $\mathbf{Q} \approx \mathbf{I}$. On the contrary, when h increases, these elements are gradually close to 1, thus $\mathbf{Q}$ is approximately an all-one matrix. During the both conditions, $\boldsymbol{\alpha}(f)$ and $\hat{f}$ keep the same with different $f_0 - f$ all the time. In the sequel, the estimation accuracy of the proposed algorithm decreases. Everything is as in Fig. 6 other than $C=1000$, Fig. 7 illustrates the MSE curves of the proposed algorithm with that h are 0.1, 1, 10 and 100, resectively. Fig. 7 is consistent with all above analyses, and we select $h=1$ in this study.

IV. Simulations and Experiments

Firstly, the impact of N on the estimation performance is considered. Everything is as in Fig. 6 except that $C=1000$, Fig. 8 illustrates the MSE curves of the proposed algorithm with that N is 8, 16, 32 and 64, resectively. It is shown, MSE performance is improved as N increases.

Next, the impact of θ on the estimation performance is considered. Everything is as in Fig. 6 except that $C=1000$, Fig. 9 illustrates the MSE curves of the proposed algorithm with that θ are -0.75π、-0.25π、0、0.5π, respectively. It is shown, MSE performances of different θ are almost the same.

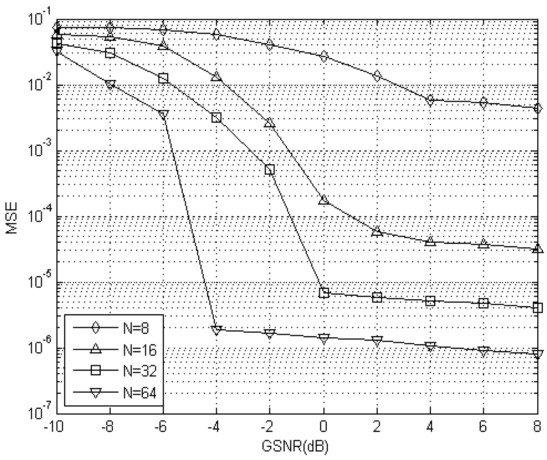


Figure 8. Impact of N on MSE

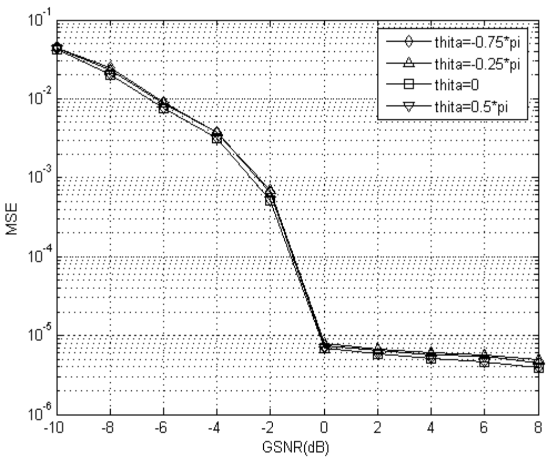
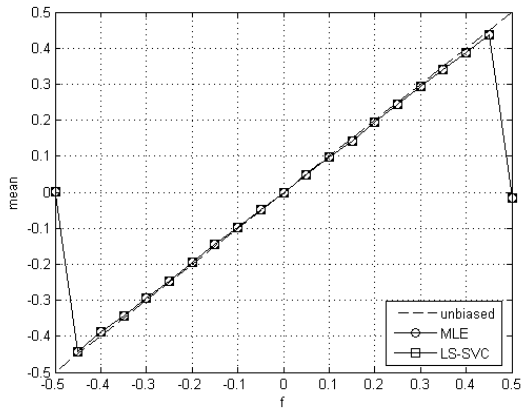
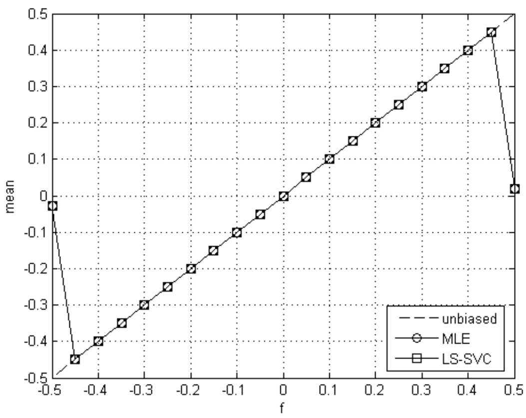


Figure 9. Impact of θ on MSE

Then, the mean performance is worked out. Under the assumption of Gaussian noise, maximum likelihood estimator (MLE) realized by fast Fourier transform (FFT) [1] is the best one because of its MSE can reach Cramer-Rao lower bound (CRLB). The proposed algorithm also searches extremum during frequency range, so it is compared with FDP, the number of FFT points is 32768. We integrate coarse and fine search in this study, whose steps are 0.005 and $1e-5$, respectively. The concrete searching method is Gauss-Newton. Everything is as in Fig. 4 and Fig. 5 except that $C=1000, h=1$, and the number of Monte Carlo experiments is 1000, the mean of both two algorithms with different f_0 are plotted in Fig. 10 and Fig. 11. Distinctly, when noise is α -stable distribution, the unbiased performance of the proposed algorithm is better than MLE one's in the condition of low SNR, and they are almost the same under other conditions.

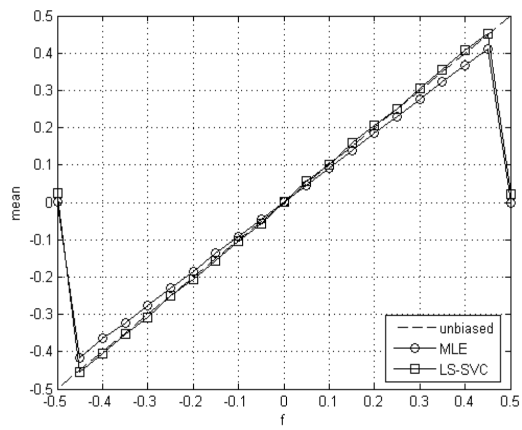


(a) SNR is -4dB

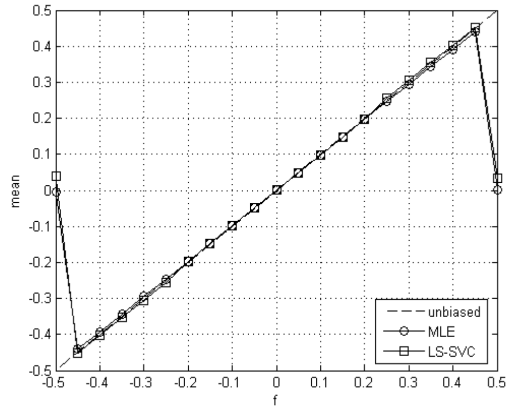


(b) SNR is 0dB

Figure 10. Mean when noise is Gaussian distribution



(a) GSNR is -4dB



(b) GSNR is 0dB

Figure 11. Mean when noise is α -stable distribution

Last, the estimation performance is taken into account. Everything is as in Fig. 10 and Fig. 11 except that f = 0.1, the MSE curves of both two algorithms are plotted in Fig. 12 and Fig. 13, respectively.

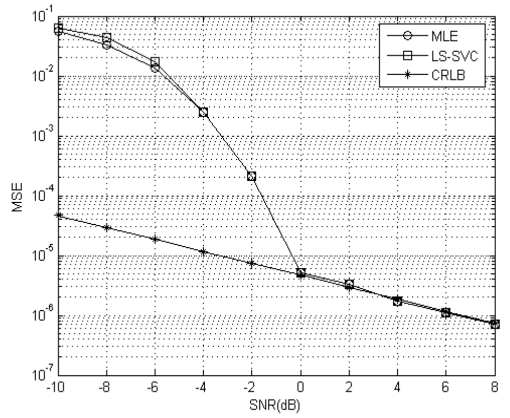


Figure 12. MSE when noise is Gaussian distribution

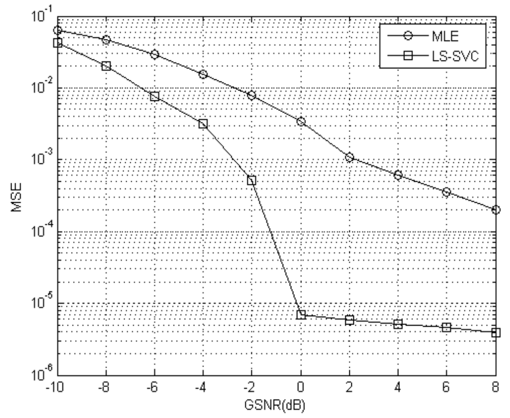


Figure 13. MSE when noise is α -stable distribution

As shown in Fig. 12 and Fig. 13, when noise is Gaussian distribution, the MSE curves of the proposed algorithm is observed to lie very close to that of MLE in the condition of known distribution model. When noise is α -stable distribution, MLE is ineffective as noise model is unknown. Nevertheless, the proposed algorithm still keep its estimation accuracy, even though that threshold effects are existing in it.

V. CONCLUSIONS

To some extent, SVC which is based on SLT can solve the problems about curse of dimensionality and over-learning. This paper converts estimation problem into classification one during the searching part, and takes use of LS-SVC to estimate frequency with unknown nois-e distribution. Also, from views of qualitative analyses and experiment results, we discuss the choice of LS-SVC's parameters. At last, we verify the feasibility and validity through simulations. In this paper, BPSK is taken for example. If signal type is QPSK, 8PSK or other modulation mode, binary classification is extended to multiclass problem in LS-SVC by the same way. At the same time, emphasized that, this paper takes Gaussian distribution and α -stable distribution as two cases of unknown noise model, but the proposed algorithm is not only orient to these two cases.

Classical algorithms of frequency estimation primarily try to change received signals into single-tone ones. On the contrary, the proposed algorithm completely consider-s effect of symbol information and directly estimates the frequency. Hence, aiming at single-tone signals, we can estimate frequency by means of adding known symbol sequence to received signals. It is a novel and converse thought, and will be valuable.

Resolving convex QP problem having inequality constraints will bring large computational load and time consuming to SVC. Although LS-SVC is presented, the proposed algorithm will search minimum value during whole range, unavoidably. As a result, there will be a tradeoff between estimation accuracy and computational complexity. At the same time, we only obtain the proper range of LS-SVC's parameters. And as a next step, how to select them exactly is an important research point.

REFERENCES

[1] H. Fu and P. Y. Kam, "Kalman estimation of single-tone parameters and performance comparision with map estimator," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4508-4511, 2008.

[2] R. G. Mckilliam, B. G. Quinn, I. V. L. Clarkson, and B. Moran, "Frequency estimation by phase unwrapping," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 2953-2963, 2010.

[3] D. Kundu, *et al*., "Super efficient frequency estimation," *Journal of Statistical Planning and Inference*, vol. 141, no. 8, pp. 2576–2588, 2011.

[4] T. Raïssi, "Reliable amplitude and frequency estimation for biased and noisy signals," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 11, pp. 4153–4157, 2011.

[5] Y. Cao and G. Wei, "A noniterative frequency estimator with rational combination of three spectrum lines," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 5065-5070, 2011.

[6] G. Wei, Y. Cao, and F. J. Chen, "Closed-Form frequency estimator based on narrow-band approximation under

noisy environment," *Signal Processing*, vol. 91, pp. 841-851, 2011.

[7] Y. Cao, G. Wei, and F. J. Chen, "A Closed-Form expanded autocorrelation method for frequency estimation of a sinusoid," *Signal Processing*, vol. 92, pp. 885-892, 2012.

[8] H. Fu and P. Y. Kam, "Phase-Based, Timing-Domain estimation of the frequency and phase of a single sinusoid in AWGN—the role and applications of the addition observation phase noise model," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 3175-3188, 2013.

[9] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*, New York: Springer-Verlag, 1988.

[10] X. Y. Ma and C. L. Nikias, "Joint estimation of time delay and frequency delay in impulsive noise using fractional lower order statistics," *IEEE Trans on Signal Processing*, vol. 44, no. 11, pp. 2669-2687, Nov. 1996.

[11] K. Barman and M. T. Arvind, "A minimum variance single frequency estimator using recursive least squares estimate of noise statistics," in *Proc. 1998 Midwest Symposium on Circuits and Systems*, Notre Dame, France, 1998.

[12] S. T. Li, H. Wan, Y. Q. Huang, and W. J. Huo, "Precise measurement of RCF in colored gaussian noise," in *Proc. 2007 International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, China, 2007.

[13] S. P. Liu, X. Wen, and Z. G. Wang, "Least square single-tone frequency estimator in non-Gaussian colored noise," *Journal of Communications*, vol. 29, no. 7, pp. 866-869, July 2008.

[14] V. N. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.

[15] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, New York: Chapman & Hall, 1994.

Xueqian Liu received his B.S. and M.S. degree both from Zhengzhou Information Science and Technology Institute in signal processing in 2006 and information hiding in 2009, respectively. He is currently pursuing the Ph.D. degree in communication engineering at Zhengzhou Information Science and Technology Institute. His research interests include digital signal processing, image processing and information hiding.